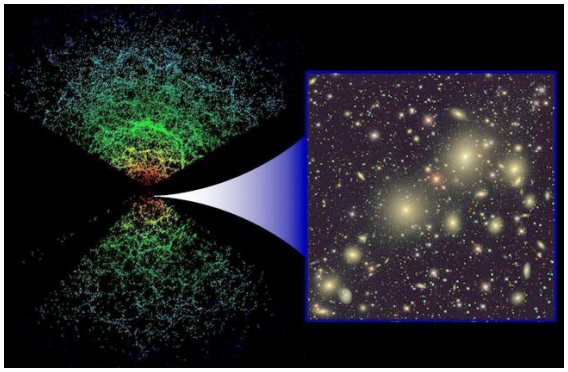
Astrophysics Computing

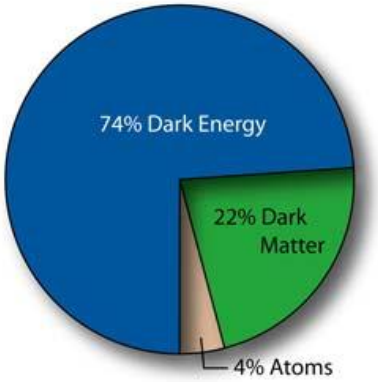
Simulations

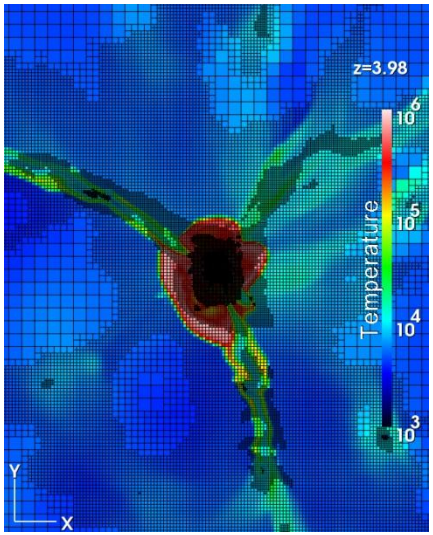
Fermilab Data Center (courtesy of Erik Gottschalk)

Need Simulations to Extract Science from Surveys

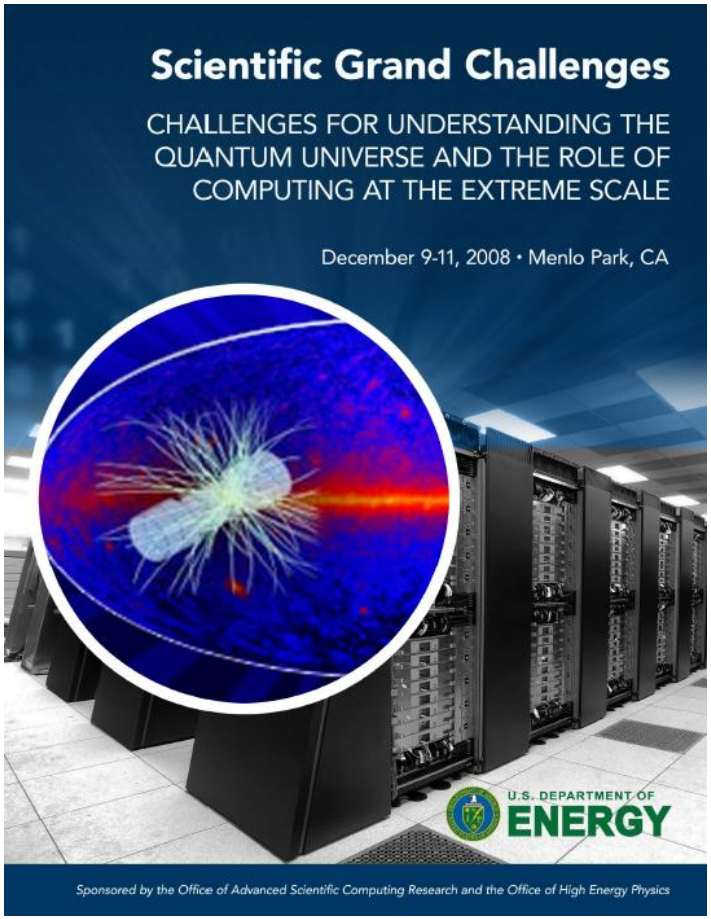


- **Dark Matter**
- **Dark Energy**
- **Neutrino Mass**
- **Inflation**





- Gravitational Instability is nonlinear
- Baryons governed by hydrodynamics
- Radiation Field affects and is affected by structure
- Stars form, Supernovae explode, …



The panel on astrophysics and cosmology simulations has identified four scientific grand challenges of direct relevance to the Quantum Universe program for which extreme scale computing will be transformational. In priority order, they are:

- cosmic structure formation probes of the dark universe
- understanding and calibrating supernovae as probes of dark energy
- particle accelerators in extreme environments
- neutrinos and the universe.

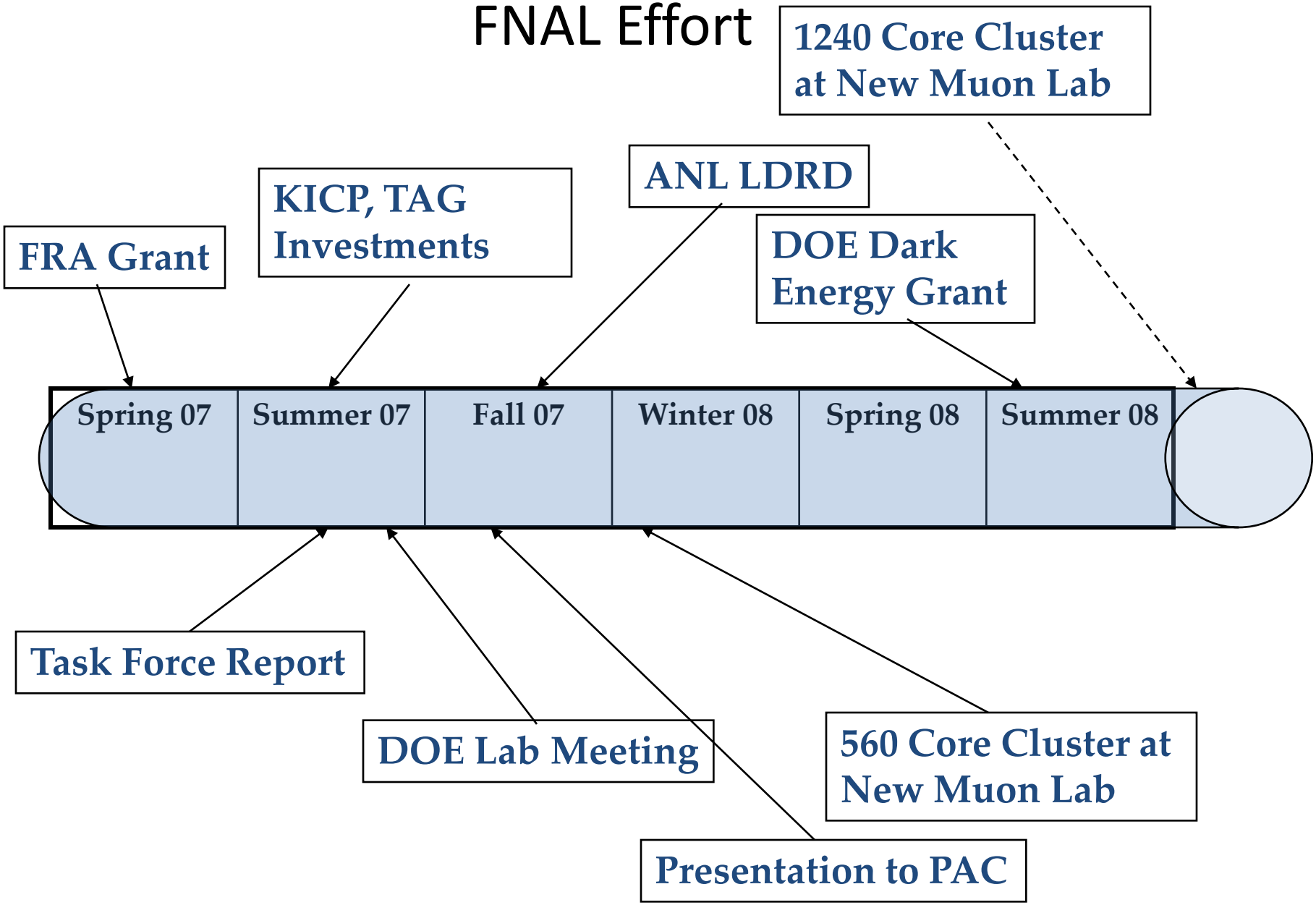
Threads are in place for Chicagoland to assume leadership role

Computational Cosmology Collaboration (C^3): Proposal

Need for Medium Scale Computing

- Cosmological surveys often require many medium-sized simulations that can be run only on a dedicated cluster
- Medium scale local clusters are ideally matched to the problems at hand
- Medium scale clusters allow further development to scale to the largest machines available
- Medium scale machines needed for analysis

FNAL Effort


1240 Core Cluster at New Muon Lab
ANL LDRD
KICP, TAG Investments
FRA Grant
DOE Dark Energy Grant
Spring 07
Summer 07
Fall 07
Winter 08
Spring 08
Summer 08
Task Force Report
DOE Lab Meeting
560 Core Cluster at New Muon Lab
Presentation to PAC

Cluster has been very successful

- Main Science machine for KICP and Fermilab theoretical astrophysics
- Dozens of papers have used the cluster
- How is it supported/upgraded?

Collaboration proposes to:

• Install and maintain **medium-scale computational facilities** enabling (a) continued development and optimization of cosmological codes for petascale machines, (b) applied computations that require fast turnaround, (c) interactive data analysis and visualization, and (d) analysis of the largest data sets produced at the leadership class facilities.

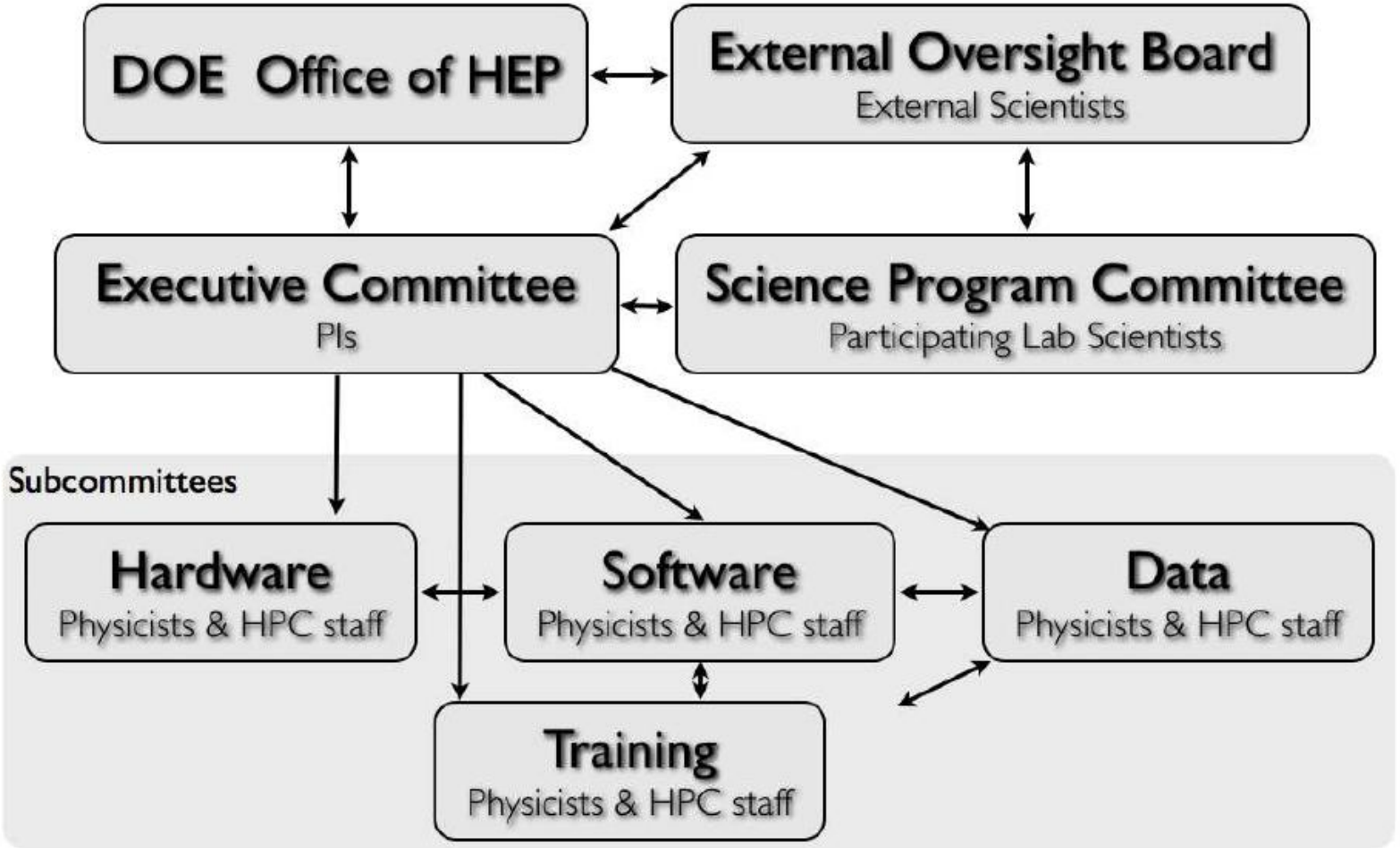
• Deliver the **Cosmology Data Grid**, a data repository of curated numerical simulations using standardized data formats, derived data products, and observational data. This user facility will be the main avenue to share and disseminate simulation and observational data within and outside the collaboration.

• Support **Computing Professionals**, whose responsibilities will include: (a) training and user support for the public software developed by the collaboration, (b) continued performance testing and supporting ongoing research to scale codes for leadership class computers, and (c) supporting the development of the cosmological data grid including curating the data products generated and organized by the collaboration.

Example: Science Driving the Request

- Dark Energy is probed by Weak Lensing
- Simulations needed to calibrate the effect of baryons on signal
- Require 3 large simulations on leadership class machines (200M CPU hours total)
- Analyses of these + many medium-sized simulations → (60M CPU hours) on local clusters
- Add up simulations required for 5 sets of cosmological probes → >100 million CPU hours/year (~10,000 cores)
- ~1 PByte of Storage & Distribution
- ~$2M/year (not including people)

Organization



Lingering Issue: Institutional Split

- 80% @ SLAC
- Medium-scale clusters @ FNAL & SLAC; large simulations run at NERSC
- Do everything @ NERSC (a la Planck)
- Start a bidding war: whichever Lab gives us the best deal, go there
- …

Challenge: Theory Culture

Physics Frontier Center Renewal

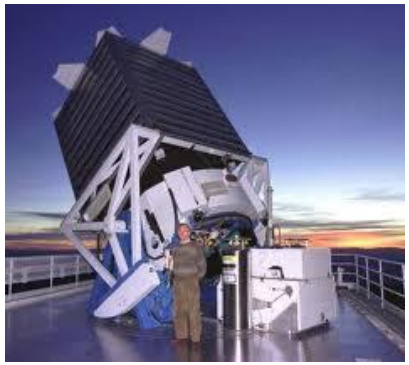
- Key Major Activity is Cosmological Computing
- Personnel (co-I's, Collaborators) include: Kravtsov, Dodelson, Gnedin, Habib, Heitmann
- Eager to Leverage the strengths of the Labs to improve the proposal
- Aim to leverage CI/PADS to develop storage, access and analysis of survey and simulation data sets

Developments at Argonne

Blue Gene/Q (2012) even better suited to local codes (more memory/node)

Habib & Heitmann (collaborators of ours) will join ANL in May (w/ 6M CPU hours on BG/P and 150M CPU hours on BG/Q)

Fermilab Data Center

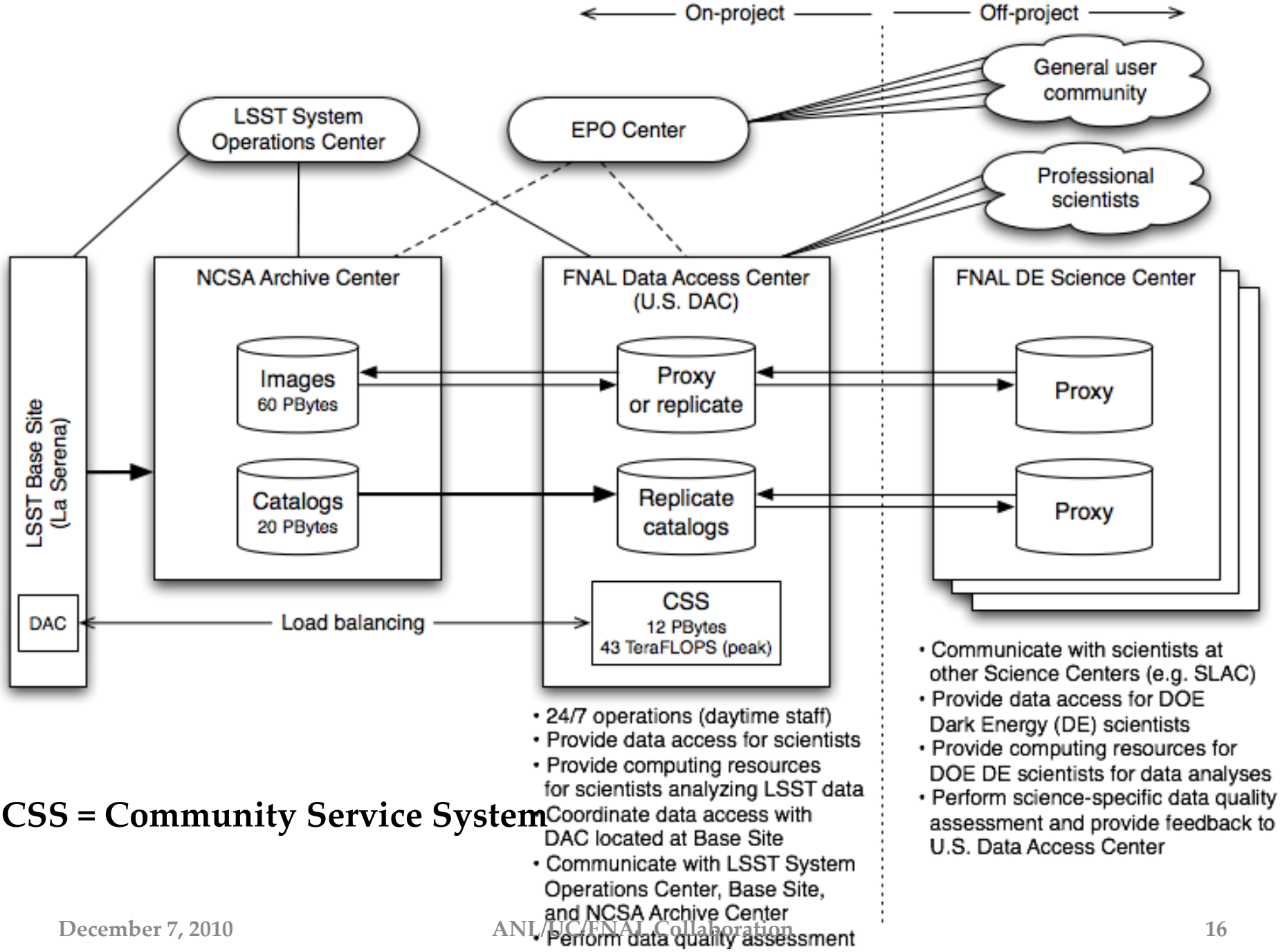


Dark Energy Survey

LSST
Large Synoptic Survey Telescope

Fermilab: an ideal site for a U.S. Data Access Center (DAC) for LSST

- Experience providing data access for the SDSS archive
- Astronomers with experience analyzing data and quality of data for SDSS
- Secondary Archive for DES
- Experience operating a 24/7 data center (monitoring data transfers and providing computing resources) for CMS Tier 1
- Fermilab pioneered and established remote operations capabilities for LHC and CMS.
- High-bandwidth network connection to NCSA Archive Center



CSS = Community Service System

Another Possible Role for Fermilab

Data Analysis Toolkit

Develop a framework and user interface for LSST science analyses. This effort would be based on the pipeline execution framework developed by NCSA for production pipelines, and involve the development of an easy-to-use analysis framework for scientists.

- The task involves development of software tools for Level 3 data products. These are the tools that scientists need to run their own pipelines for science analyses.
- The technical challenge is to take software infrastructure developed by others (framework, middleware, database access, etc.) for production pipelines and make it easy for scientists to use for their own analyses.